EuroCALL 2025: Advancing CALL: New research agendas
27-30 August 2025, Università Cattolica del Sacro Cuore, Milan, Italy
DOI: https://doi.org/10.4995/EuroCALL2025.2025.21217

AI-enhanced graded viewers: Creating level-appropriate video content for language learning

Euan Bonner
Center for Learning and Teaching Innovation, Kanda University of International Studies, , euan.bonner@gmail.com

How to cite: Bonner, E. (2025). AI-enhanced graded viewers: Creating level-appropriate video content for language learning. In Y. Choubsaz, P. Díez-Arcón, A. Gimeno-Sanz, V. Morgana, A. C. Murphy & F. L. Seracini (Eds.), *Advancing CALL: New research agendas - EUROCALL 2025 Short Papers.* https://doi.org/10.4995/EuroCALL2025.2025.21217

Abstract

Just as graded readers provide language learners with level-appropriate texts to develop reading skills, 'graded viewers' can deliver customised audiovisual content for targeted listening practice. The widespread use of language learner focused video content however is often hindered by a limited selection and a lack of engaging, level-appropriate content, due to its significant cost of production, limiting effective self-study. To address this gap, the author has developed a proof-of-concept AI-powered 'graded viewer' system that generates short customisable television-style videos and auto-graded comprehension quizzes. The system integrates Large Language Models (LLMs) for script generation, real-time 3D animation for visuals, and neural voice synthesis for dialogue, allowing users to create content tailored to their CEFR level across five different genres. Preliminary feedback from 20 student users indicates that the system successfully generates comprehensible input and its level of customisation is highly engaging. However, challenges in maintaining narrative coherence and ensuring consistent quality control remain. This on-going project seeks to demonstrate the potential for AI-generated graded viewers as a new category of autonomous, user-driven language learning resources that have the potential to bridge the gap between engaging authentic media and pedagogically sound educational materials.

Keywords: *artificial intelligence; educational video; personalized learning; comprehensible input.*

1. Introduction

In language learning, the use of authentic video content is a key component in giving language learners exposure to natural speech, cultural nuances, and contextualised language use (Gilmore, 2007; Weyers, 1999). However, the effective integration of this content into a student's self-study regimen is often hampered by a fundamental challenge: The difficulty of sourcing materials that are simultaneously engaging while also being comprehensible and linguistically appropriate for the learner's proficiency level. As a result, learners often face a trade-off; they can choose engaging, entertainment-focused content that is not specifically designed for their needs and is almost certainly linguistically overwhelming, or select from a very limited range of pedagogically sound materials that often lack narrative appeal. This results in a compounded issue of having a lack of entertaining content to maintain motivation, and a scarcity of level-appropriate content across specialised domains (e.g. business negotiations & flight services).

Hence students are much more likely to engage with the vast variety of 'graded readers' that have been developed for language learning. These books and audiobook materials provide level appropriate text in many specialised domains while also providing additional features such as glossaries and quizzes to aid students in checking their

comprehension. While audiovisual graded 'viewers' do exist, language learner-focused video materials are extremely limited in range, primarily due to the costs involved in their creation compared to more traditional graded reader and audiobook materials.

However, the advent of large language model (LLM) artificial intelligence (AI) presents a novel opportunity to overcome these production cost barriers. There is great potential for AI-generated content to bridge the gap for language learners, aiding them in experiencing level-appropriate video content that is dynamic and customisable in a way that matches their specific language learning needs. As a result, the author has developed an AI graded viewers application to assess the current capabilities of LLMs in their ability to address this gap.

This application features a novel system that leverages recent advancements in LLMs to generate customisable graded viewer content. The system integrates three core technologies: LLMs for on-demand script generation and scene direction, real-time 3D character animation within dynamic environments, and high-quality natural neural voice synthesis to bring the character voices to life. The goal is to empower learners to become active co-creators of their learning materials. By providing customisable parameters such as Common European Framework of Reference for Languages (CEFR) proficiency levels, topics of personal interest, specific target vocabulary, speaking speed, and desired content genres, students can generate unique, short-form television-style episodes that have never been seen before.

The purpose of this paper is to provide a detailed description of this application and is presented as a work-in-progress report on its current status. It will first outline the application's core capabilities, the system's architecture, the user's workflow, and the design of its key features. Following the system description, the paper will discuss some of the pedagogical principles that informed its design and conclude with a discussion of its current limitations and directions for future development as the project moves closer to completion.

2. The AI graded viewer application

2.1. Background

Researchers have explored the potential of LLMs for generating customised reading materials for language learners, aiding them in creating level appropriate texts in a dynamic and customisable way that matches individual student needs (Bonner, et al., 2023; Godwin-Jones, 2024). However, the extension to video content remains largely unexplored. The primary barrier has been production cost and complexity. Creating even a single educational video requires scripts, actors, filming, editing, and post-production, making it financially prohibitive to produce the volume and variety needed for comprehensive language learning on the same scale as graded readers. AI-generated video content could overcome these barriers by automating the entire production pipeline, from script writing to voice synthesis to visual presentation.

The raw generation of visual and audio content to create what is now commonly referred to as AI-generated video is a rapidly evolving field. After a few years of rapid improvements, in 2025 Google's Veo 3 AI video generator became the first to generate near lifelike animations in any genre accompanied by lip-synced voices and sound (https://gemini.google/overview/video-generation/). However, due to current technological limitations, these videos often remain exceedingly short and hard to control. Veo 3, for example, launched with videos limited to 8 seconds, and is challenging for creators to guide to produce the specific scenes and content they wish to see and hear, necessitating multiple attempts that each take minutes to generate. The generation process can be time-consuming and is currently more akin to a lottery in terms of creating content that matches the user's specific desires.

Until this technology matures, however, an alternative approach to AI-generated video content may offer a more practical solution in the interim. There is great potential in combining a number of roles that LLMs can fulfil to create bespoke video content that meet the needs of content creators and viewers. AI can simultaneously act as a scriptwriter, a director, a sound engineer and as actors. When combined with live 3D animation, there is potential for it to be taught how to create videos in their entirety, managing everything necessary to meet the needs of viewers' desires. Several projects have already begun exploring this hybrid approach to AI-generated video content.

Since 2022, the "*Nothing, Forever*" show on Twitch (https://www.twitch.tv/watchmeforever) has been broadcasting an almost non-stop AI-generated parody of the 1990s sitcom Seinfeld (Cantor, 2023). Utilising 3D sets and character models, the cast engages in endless comedic conversations that embody the show's mantra of being a show about nothing. Similarly, companies like Fable Studio (https://www.showrunner.xyz/) are developing their "Showrunner" service to create entire cartoons based on user input. While not as dynamic as a human-generated cartoon or live-action TV show, this kind of hybrid AI-generated video content can potentially bridge the gap between the lack of language-focused video content available for learners and the eventual arrival of fully AI-generated animation. Hence, this project was formed to investigate the creation and utilisation of a similar hybrid computer/AI-generated content creation platform. In contrast to these entertainment-focused platforms, this system is designed specifically for educational language learning uses. While the technical side of this application draws inspiration from these entertainment platforms, the educational focus of this system necessitated careful consideration of language learning pedagogy in its design.

2.2. Pedagogical design and justification

Each feature was designed based on specific pedagogical principles aimed at enhancing independent language acquisition. The application's primary pedagogical goal is to provide learners with a limitless stream of comprehensible input (Krashen, 1982). Influenced by the principles of extensive reading, the application has been designed to enable students to tap into an unlimited source of rapidly-generated easy-to-understand user-generated content, ensuring that they have extensive exposure to vast amounts of input, which has been shown to increase fluency and motivation (Nation, 1997).

In the same manner that graded readers are ‘graded’ according to each publisher’s CEFR level equivalent system, the first step of the application is to choose their desired level. This generates content following Krashen’s (1982) 'i+1' concept, which posits that learners acquire language "by understanding input that contains structures a little beyond our current level of competence" (p. 21). This ensures that the linguistic output is just challenging enough to facilitate acquisition but not be overwhelming like in a video targeted at native speakers. Content must also be compelling. Learners are more receptive to acquisition when they find content relevant to their interest, as the most effective input should be so engaging that the acquirer "may even 'forget' that the message is encoded in a foreign language" (Krashen, 1982, p. 74).

Hence the need for features that allow users to select personally interesting topics and genres. Furthermore, the use of the 3D animated environments serves as a comprehension scaffold. Within the context of extensive listening, multimedia provides rich visual support that, as Vo (2013) notes, “helps to promote students' listening comprehension and is more facilitative for less proficient language learners” (p. 31). The characters, setting and basic onscreen actions in the application serve as a persistent visual context that helps ground the dialogue and make the language more accessible. Furthermore, interactive features like adjustable playback speed and script viewing provide learners with additional real-time scaffolding tools to manage comprehension.

While the application does feature the capability for users to create their own shows from scratch, the system was designed to avoid the classic paradox of choice (Schwartz, 2004), wherein too many options give rise to decision paralysis. Learner performance can decrease when the number of choices becomes too large, hence the importance of assembling a small curated set of options via the pre-set genres feature. Schneider (2021) states that there needs to be a 'sufficient guideline to foster learning retention and transfer processes,' which is a 'minimum of three to five choice options' (p. 15). Therefore, the decision was made to offer five preset shows across different genres, allowing users to avoid overthinking their content generation. These pedagogical considerations informed the technical design and implementation of the system's architecture.

2.3. System components

The application’s architecture consists of several local and cloud-based systems. The core of the application runs on the Unity engine (https://www.unity.com/), and was designed to be compatible with locally run (e.g. Meta Llama, DeepSeek etc.) and major cloud-based LLM services (e.g. OpenAI’s GPT range, Anthropic’s Claude, Google Gemini, etc.) for script generation. This allowed for the testing and comparison of all currently available LLMs during the development period. Similarly, several cloud-based neural voice synthesis services were integrated (e.g. Microsoft Azure, Google Cloud Voices, OpenAI Voice Agents etc.), and Runware’s Flux Schnell

(https://runware.ai/) was used in order to provide AI-generated episode title cards that could be generated and displayed within 2-3 seconds of user submission. Graded viewer learner assessment is handled by a linked external web-based quiz system also developed for this project.

2.4. The user workflow

Upon launching the application, users create an account or log in via a five-digit code. The main menu (see Figure 1) offers quick-start mode, previous episode loading, or custom show creation. Users select their CEFR level (A1-C2), which adjusts vocabulary complexity, grammar, and speaking speed. They input a topic (from a single word to a full sentence) and optionally add target vocabulary.

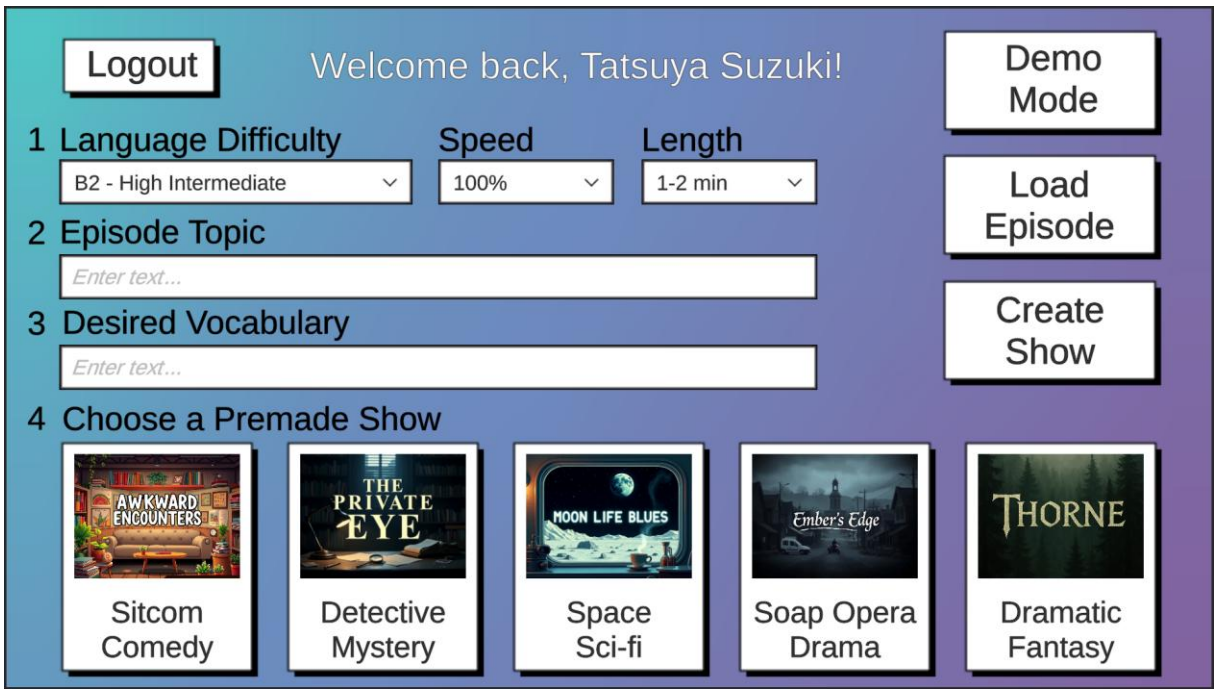

Figure 1. The main menu of the application, showing the setup process.

Users then choose from five preset shows spanning different TV genres: *Awkward Encounters* (sitcom comedy), *The Private Eye* (mystery), *Moon Life Blues* (sci-fi), *Ember's Edge* (soap opera), and *Thorne* (fantasy). After selection, a brief static screen displays while the AI generates the script and title card. Episodes feature 2-4 characters in genre-appropriate locations and clothing, with characters automatically navigating and interacting with their environment and each other while delivering AI-generated dialogue (see Figure 2). During playback, users can view the full script, pause, or adjust vocal speed in real-time. After the episode concludes, users can replay episodes with enhanced line-by-line navigation controls. The system also generates a multiple-choice quiz and uploads it to a web-based hub accessible via QR code on smart devices. The online hub stores previous quizzes, scripts, scores and which quizzes have not yet been taken.

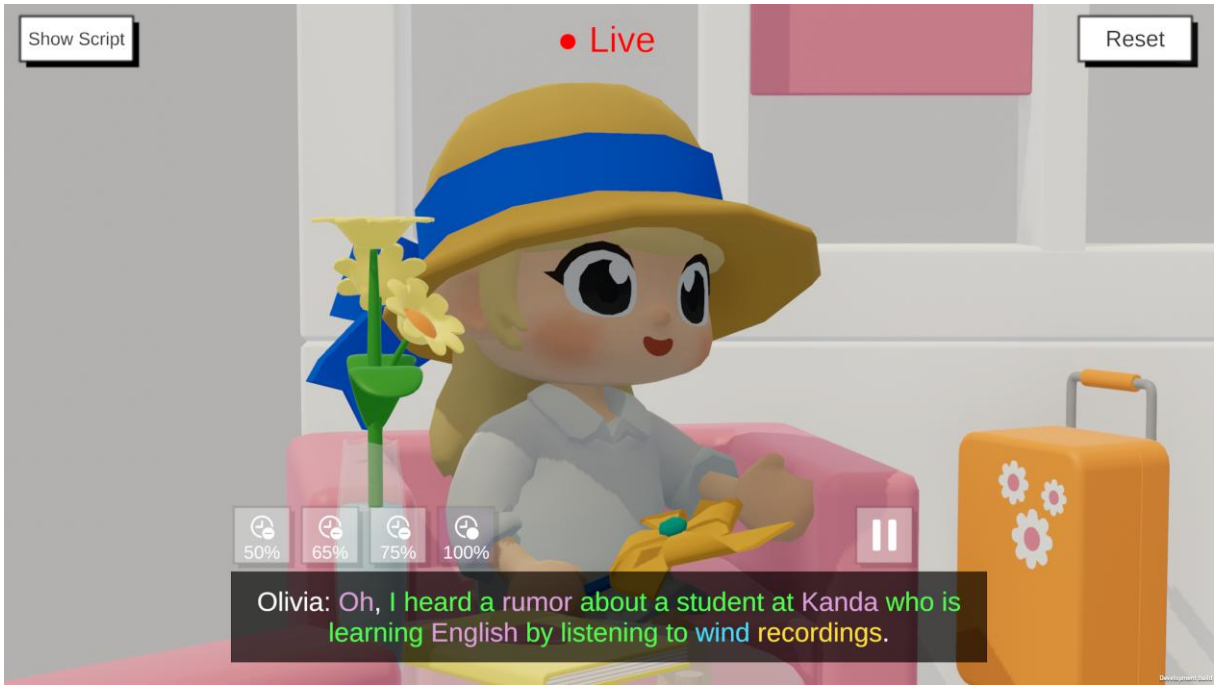

Figure 2. A scene from the soap opera drama genre discussing "weird ways to learn English." Subtitle colour denotes English word frequency.

2.5. The content generation process

This streamlined user experience masks a sophisticated backend process where multiple AI systems work in coordination to generate the educational content. Behind the scenes, when users submit their episode parameters and select one of the premade genre-specific shows, the system initiates a multistage content generation process. First, the application pieces together various LLM prompt elements and combines them together into a 'master prompt' that will be submitted to the LLM. These prompt elements consist of specific instructions related to the CEFR level, the desired topic, chosen vocabulary, and genre-specific TV show preset information. Simultaneously, a separate image generation LLM creates a television show title card based on the user's chosen topic and episode genre (see Figure 3). This process uses a genre-specific prompt to ensure the generated image matches both standard title card conventions and the selected show format. Until the image returns from the image generation LLM, a static noise animation is played on the screen mimicking the retro-style tuning of an old television. Once the title card appears, the genre specific music starts to play until the script LLM returns with its properly formatted episode content. The application then loads the LLM's desired set, and generates the characters. The whole process typically takes less than ten seconds. Subsequently the first line of dialogue is parsed from the script, identifying which speaker it belongs to. Based on the character profile, the line of dialogue is submitted to the voice synthesis service along with the gender, accent and speaking tone of that character. Once this first line of synthesized dialogue is returned, the title card fades out and the TV show begins. While the first line of dialogue plays, the next line of dialogue is being generated.

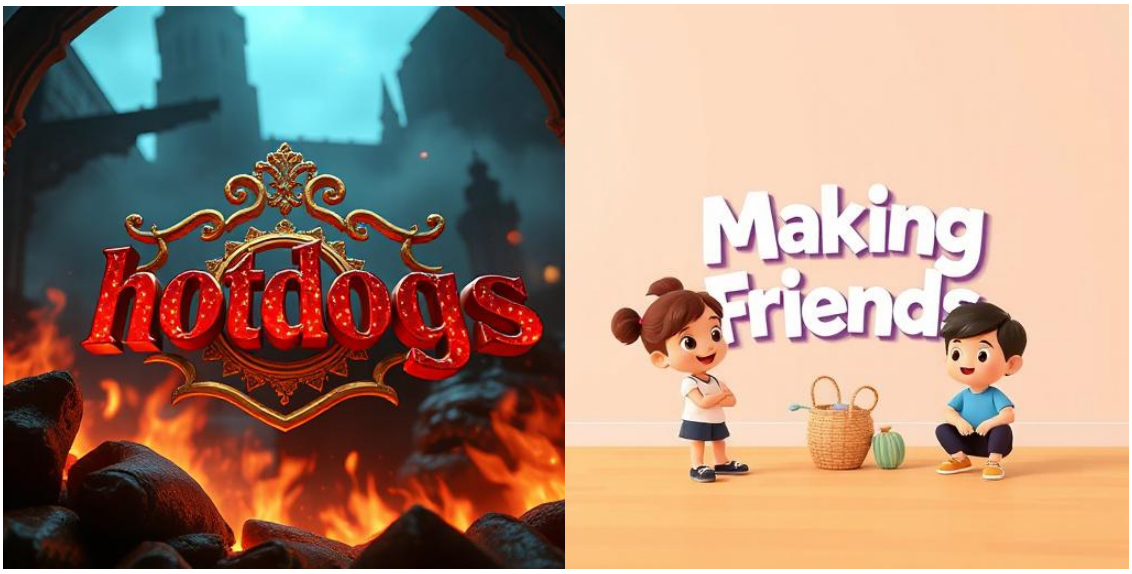

Figure 3. AI-generated television show title cards based on student topics and genres: (left) Fantasy: hotdogs & (right) Comedy: making friends.

At the conclusion of the episode, the genre-appropriate end credits music plays, and the same title card is displayed again while the system prompts the LLM to generate a CEFR-levelled multiple choice quiz based on the script with answers and feedback included. Once that quiz arrives, it is sent to an online website designed for this project that hosts and formats the quiz and returns a QR code that is displayed onscreen. The user is then able to scan this code and take the quiz on their smartphone. When they submit the quiz answers they are automatically provided with a score, answers, and feedback. The script and quiz are added to their online account for later review.

2.6. The genre-specific prompting system

To ensure quality and consistency of the generated content, the prompts submitted to the script generator LLM are carefully engineered to ensure that the LLM only takes creative liberty in the specific way desired in order to create a new and interesting episode of the existing preset TV shows. Each of the five genres has a unique ‘master prompt’ that provides the LLM with an extensive set of instructions about what kind of script it is generating, its specific formatting, real world examples from which it can draw inspiration, and the show’s overall premise, character profiles and backgrounds. As part of the script generation process, the LLM is also instructed to choose from a list of possible character appearances, and a list of available locations for that genre for the scene to take place within. These genre specific instructions are crucial for guiding the LLM beyond simply generating conversations to generating specific, structured, thematic content typical of a television show of that genre.

3. Early implementation and user experience

While this project is still currently underway with a more extensive data collection soon to be conducted, feedback from some initial users present some interesting observations.

3.1. Study design

To gather preliminary data on the application’s usability and user experience that would inform further improvements before the main data collection period, a pilot study was conducted with university-level English language learners at a communications-focused university in Japan. The study received ethical approval from the university's institutional review board and consisted of 20 students taking part in one-hour sessions consisting of a five-minute orientation, thirty minutes hands-on period and then an open-ended survey. Participants' English proficiency levels ranged from TOEFL 440-550 (approximately A2/B1 to B1/B2 on the CEFR scale). During the hands-on period, students created, watched, and took quizzes for anywhere from three to ten TV show episodes each. All students were able to successfully follow the instructions and take the quizzes without issue, suggesting

that the application is both accessible and at least somewhat compelling for the learners. The survey conducted afterwards consisted of three basic open-ended questions asking for their feedback in the following areas:

- TV-shows' stories, dialogue, and language level
- Post-quiz appropriateness of questions and difficulty level
- Usefulness of the application as a language learning tool

Students were free to answer these questions in English or in their native language.

3.2. Key feedback themes

Customisation: The most prevalent theme from participant feedback, mentioned by over half of the participants (11 of 20) was the motivational power of customisation. Students consistently reported that the ability to generate videos based on their own specific interests and proficiency levels made their learning process more engaging and personal. One participant noted: "自分が興味のあるテーマやジャンルで学べるという点も、継続的な学習へのモチベーションにつながります" [The ability to learn with topics and genres I'm interested in leads to motivation for continued learning]. Comments like this suggest that the application is successful in its primary goal of empowering language learners to create their own engaging content.

Content Quality: Feedback about content quality was more nuanced. While nearly half of the participants (9 of 20) commented positively that the generated episodes were coherent and interesting, several specific challenges were highlighted. The most common was a difficulty with cultural context, especially about the content of the comedy genre. Four students reported that while they understood the content, they could not understand the humour or 'punchlines' of the jokes. As one student put it: "特にコメディジャンルでは、英語レベルの問題というよりも、笑いの「ネタ」がうまく理解できず、何が面白いポイントだったのかを十分に捉えられないこともありました。" [Especially in comedy genres, rather than a problem with my English level, I sometimes found myself unable to fully grasp what was funny about the jokes]. Others noted user supplied vocabulary would often get overused and that the AI voices could sound flat and lacked emotion in their delivery (though it should be noted that episodes aimed at A1-B1 CEFR levels utilized slower, less naturally speaking synthesized voices). Furthermore, 7 participants noted difficulty understanding higher CEFR level-generated content, though it is hard to determine whether that was a result of participant language levels or if the scripts are using overly technical vocabulary at higher levels. One student noted: "From around B2 onwards, I started to come across words I didn't know. C1 has a lot of technical terminology which I hardly understood." Overall, this feedback suggests that future work must focus on refining the LLM's understanding of cultural nuance and further investigation into its ability to stay within vocabulary range limits.

Assessment: The feedback on the quizzes stands out with clear systematic issues often found in LLM-generated assessment. Correct answer positional bias was explicitly mentioned by five participants (e.g. "If the answer of the first question is no.1, other questions' answer is also no.1!"), while length bias was identified by three others (e.g. "選択肢に必ず1つ長文があったので適当にそれを押したら全部あってた" [There was always one long sentence in the choices, so I just randomly pushed it and got them all right]). This confirms that the quiz component, while functional, will require significant refinement and assessment as to whether or not it can serve as a valid assessment tool.

4. Quality issues and system limitations

4.1. Content quality issues

The primary limitation of the current system is the limited ability of LLMs to structure their content following a narrative path consistent with short format television content. This is an issue that is rooted in their nature as sequential text predictors, rather than forward thinking, narrative arc following plot planners. A primary challenge in the current system is maintaining narrative coherence, particularly across scene and location transitions. Early in the process of creating the application, there was a desire to create multi-scene scripts that would transition between various locations. However, the LLM often struggles to build upon previously generated content when a scene changes, sometimes treating the new scene as the beginning of a new unrelated episode or only considering the reason for the transition after the fact, creating unnatural convoluted reasons for the characters to continue their

previous conversation within the new location. This issue stems from the inherent nature of LLMs. As sequential text predictors, they do not plan long-term narrative arcs well. They are designed instead to tackle each line of the script based on the genre's expected structure. For example in terms of a comedy show, structure dictates that they need to conclude with some kind of punchline that brings everything together, but the LLM will not consider that punchline during the creation of the earlier scenes of the episode, instead waiting until the last moment to look back upon what they have generated so far and try to come up with some kind of punchline then and there, unable to edit what they have already written. This can be somewhat mitigated by instructing the LLM to first generate a detailed episode synopsis before writing the script, providing it with a roadmap to follow during dialogue generation. However, this approach merely displaces the coherence problem to a higher level of abstraction, the LLM faces the same sequential prediction limitations when creating the synopsis itself, still unable to plan backwards from a satisfying conclusion or consider how early story elements should be structured to support later narrative payoffs. The fundamental challenge of creating seamless, narratively motivated content remains an area for future development and refinement.

Content quality control issues also arise from a mismatch between the topic and the chosen genre. Users are free to write whatever topic they wish and then choose whichever genre they desire, and typically the LLM can generate a script that is able to blend these two considerations into an entertaining episode. Of course, there are instances where the topic and the genre are almost incompatible (e.g. Topic: Working at a call centre, Genre: Epic Fantasy), however this typically results in an entertaining mismatch rather than a lacklustre or confusingly written one. The issue itself is that on occasions the LLM, when presented with a mismatching topic for the genre, will sometimes elect to ignore the topic and generate its own topic of conversation that matches the genre better. This can result in users feeling a sense of frustration that the AI did not take their ideas into consideration. This problem has been mostly mitigated through extensive prompting, but as with all things solved solely with prompt adjustments, LLMs are prone to simply ignoring these instructions on occasion.

Another issue is tied to the system's reliability in adhering to CEFR levels when generating content. Selecting lower levels such as A1-B1 does consistently produce simpler sentence structures accompanied with basic vocabulary and grammar, however the LLM tends to over focus on producing high-level language content at the C1-C2 level. Dialogue produced at these levels often becomes excessively verbose, with the dialogue aiming to utilise unnecessarily complex synonyms wherever possible rather than adhering to specific CEFR level appropriate vocabulary and grammar structures. It has been observed that more recent improvements to LLMs do this less often, however adhering to specific vocabulary lists and grammar points remains a 'soft' constraint rather than a 'hard' rule, with the LLM still occasionally deviating into flowery verbose language. It should also be noted that this difficulty, however, may not be solely a limitation of the technology but also a reflection of the inherent ambiguity in the CEFR framework itself, as human educators often face similar challenges in consistently operationalizing its high-level descriptors.

4.2. Assessment reliability issues

User feedback so far has confirmed that the AI-generated episode quizzes suffer from the known issues of generating multiple choice assessment using LLMs. The auto-generated multiple-choice questions suffer from several issues common in current LLM-based quiz design: There is significant positional and strength bias when generating multiple choice questions. The correct answer is frequently placed in the same response position (e.g. option B) for each question, and is conspicuously longer or more detailed than the incorrect option distracters. Additionally, creating prompts that consistently produce questions of an appropriate difficulty (and avoid questions that are either trivial or require expert knowledge beyond what is provided in the script) remains a significant challenge. This kind of issue suggests that while the quiz component of the application serves as a conceptually useful check of student understanding, significant refinement is needed before this tool could be considered reliable for formal learner assessment in a similar way to that provided in graded readers. It may also be worth reflecting on whether multiple choice questions are the best method of assessment, and future development should also include exploring alternatives, such as providing open-ended comprehension questions or other forms of non-graded, formative feedback that encourage reflection over performance.

4.3. Technical implementation constraints

Due in part to the current limitations of what LLMs are capable of in terms of scene direction, direct character control, the defined scope of this project and the time needed to program its desired features, the character animations are not directly tied to the script's content. Characters play speaking and gesturing animations when it is their turn to talk, look at the previous speaker and turn their heads towards the current speaker when listening. However, their interactions with each other and the location they are within are based on a separate AI driven navigation system. This results in characters walking, sitting, eating, and using items such as computers regardless of their level of participation in the conversation. As a result of this limitation, each of the preset genres mimic shows where dialogue drives the entertainment, rather than physical action. As an example, sitcoms such as Seinfeld derive their entertainment value from what the characters are saying and not what they are doing. Whether they are in a cafe eating lunch, at home pouring a bowl of cereal, or standing outside on a busy street, often the topic of conversation has very little to no relation to where they are and what each character is doing while they speak. Therefore, the application's preset shows were deliberately designed to mimic this dialogue-driven style, where the entertainment value is derived primarily from the situated linguistic content rather than physical action.

5. Future directions and conclusion

5.1. Future directions

The next stage of the project is to take these limitations and user feedback and continue to develop the application. Future development will focus on directly addressing the above limitations as much as is possible in order to determine whether LLM-generated language learning focused video content is a current capability of LLMs or if it remains a potential future affordance. Key priorities include:

- **Refining the Quiz System:** Positional bias can be solved by simply randomizing the order generated by the LLM, and length bias may be solvable through prompting in similar word length requirements for all multiple-choice options.
- **Improving Narrative Prompting:** Experimenting with more advanced prompting techniques and utilizing the latest LLMs as they become available may help improve multi-scene and full episode narrative coherence.
- **Enhancing Pedagogical Scaffolding:** Incorporating user-requested features such as an integrated dictionary could allow students to look up words simply by touching them in the subtitles.
- **Adhering to Cultural Expectations:** Current shows have a bias towards US/UK television genre conventions. By knowing more about the user before generating content, the LLM may be able to find a better balance between learner expectations of genres and exposing them to genre conventions of a specific target culture.

While other features may not be solvable in the short term, such as comedic timing and word stress issues with speech synthesis, focusing on the above will help transition the application from a successful proof-of-concept to a more robust and pedagogically reliable tool for autonomous language learning.

5.2. Conclusion

This ongoing project hints at the promise that the creation of AI-powered graded viewers for language learning is on the cusp of being technologically feasible. However, it also underscores that the technology exists currently in a state of promising imperfection. The system's greatest strength lies in its ability to provide truly personalised content, with users consistently reporting high engagement when learning with topics and genres of their own choosing. While it is fully capable of generating a limitless stream of customised, comprehensible input, the system's lack of reliability in areas of narrative coherence, pedagogical precision and cultural nuance highlight the remaining gaps between this functional proof-of-concept and a fully robust and reliable educational tool. Looking ahead, this project aims to move beyond a general assessment of the application's capabilities to a more granular analysis. Future work aims to produce an actionable framework focused on systematically identifying under which conditions LLMs can reliably generate pedagogically sound video content, and which areas require further research and development, both pedagogically and technologically. Ultimately it is hoped that the AI graded

viewer concept should not be seen as a replacement for existing tools, but as an entirely new category of extensive learning resource. It may serve as a powerful supplement for autonomous learning, uniquely bridging the gap between the static, structured world of graded materials, and the near limitless, unlevelled content of authentic media. As the underlying technology of LLMs continues to mature, its potential to provide truly personalised and compelling language learning experiences will only grow.

References

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *English with Technology*, *23*(1), 23-41. https://doi.org/10.56297/BKAM1691/WIEO1749

Cantor, M. (2023, February 5). AI Seinfeld: the show about nothing is back – and now it's written by robots. *The Guardian*. https://www.theguardian.com/tv-and-radio/2023/feb/04/ai-seinfeld-nothing-forever-twitch

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, *40*(2), 97–118. http://dx.doi.org/10.1017/S0261444807004144

Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language Learning & Technology*, *28*(2), 5–31. https://hdl.handle.net/10125/73570

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press. https://www.sdkrashen.com/content/books/principles_and_practice.pdf

Nation, P. (1997). The language learning benefits of extensive reading. *The Language Teacher*, *21*(5), 13–16. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/1997-Benefits-of-ER.pdf

Vo, Y. (2013). Developing extensive listening for EFL learners using internet resources. *TESOL Working Paper Series*, *11*(1), 29-47. https://www.hpu.edu/research-publications/tesol-working-papers/2013/02_YenVo2013.pdf

Schneider, S. (2021). Are there never too many choice options? The effect of increasing the number of choice options on learning with digital media. *Human Behavior and Emerging Technologies*, *3*(2), 307–319. https://doi.org/10.1002/hbe2.295

Schwartz, B. (2004). *The paradox of choice: Why more is less*. Harper Perennial.

Weyers, J. R. (1999). The effect of authentic video on communicative competence. *The Modern Language Journal*, *83*(3), 339-349. https://www1.wellesley.edu/sites/default/files/assets/departments/pltc/files/faculty/weyers_1999_authentic_video_and_communicative_competence.pdf

